

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-202820
(43)Date of publication of application : 04.08.1995

(51)Int.Cl.

H04B 14/04
H04N 7/30
H04N 9/802

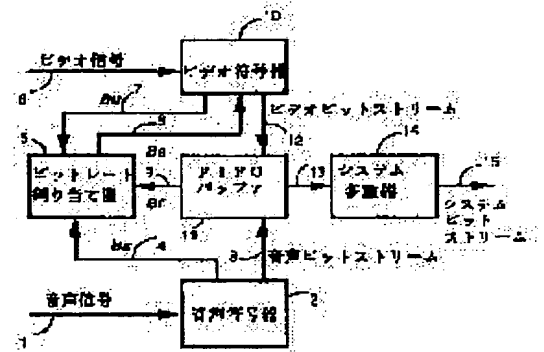
(21)Application number 05-336975 (71)Applicant : MATSUSHITA ELECTRIC IND CO LTD
(22)Date of filing : 28.12.1993 (72)Inventor : FUAN SHII YUN
TAN OO PEN

(54) BIT RATE CONTROL SYSTEM

(57)Abstract:

PURPOSE: To improve the quality of a video signal by using a mechanism controlling an image rate together with a picture coding mode so as to use available bits for a prescribed period more efficiently.

CONSTITUTION: After each audio frame at an input 1 is coded in a voice coder 2, bit saving is calculated and bit number Bs4 saved in the audio frame is calculated for a preceding image coding period based on the request of a bit rate allocating device BRA 5 and the result is fed to a BRA 5 for video coding. Simultaneously a bit total number Bu7 used for a preceding image is fed from a video coder 10 and a sufficiency degree Bf9 is fed from an FIFO 11 to the BRA 5. The image coding mode is selected by the coder 10 based on the total number Ba8 from the BRA 5. A bit quality to be consumed 7 of a current image is fed to the BRA 5 for succeeding image coding. The coded audio signal 3 and the video bit stream 12 are multiplexed by a multiplexer 14 and the audio signal and the video signal are synchronized.



LEGAL STATUS

[Date of request for examination] 01.07.1999
[Date of sending the examiner's decision of rejection] 21.03.2001
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

*** NOTICES ***

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] The Variable Bit Rate voice coder which is the bit rate control system which controls a bit rate dynamically, encodes a sound signal, and gives the saving bit total for video encoders (B s), The Variable Bit Rate video encoder which encodes an input video sequence signal per image of per-second fixed numbers, The bit rate quota machine which assigns dynamically the bit total which can be used for the present image encoded using the output of said Variable Bit Rate voice coder, The FIFO buffer which buffers the bit stream of the encoded voice and video and gives the fixed bit rate bit stream of the combined video and related voice, The bit rate control system characterized by having voice and the multiplexer which multiplexes the bit stream of video by the time stamp added in order to synchronize the signal of voice and video.

[Claim 2] Said Variable Bit Rate voice coder is coordinated with voice prediction analysis. A voice frame A voice detection / discernment means to judge in advance whether the suitable bit rate which needs to be used for coding of a voice frame is needed, or it may be skipped (ADI), The bit rate control system according to claim 1 characterized by having a sum total voice bit rate count means to calculate the saved bit total (B s) and to send it to a bit rate quota machine (BRA) during a precedence image coding period.

[Claim 3] In order to search for distribution of energy level and energy, said voice detection / discernment means (ADI) A sound signal class mainly in a silent period from a means to analyze the frequency and energy component of an input sound signal, and the aforementioned energy level and distribution of energy A non-vocal sound signal or a means to mainly classify into a vocal sound signal, The bit rate control system according to claim 2 characterized by having the coding inaccuracy permitted and a means to predict a quantization noise so that a predetermined bit rate may be produced for voice coding.

[Claim 4] said bit rate quota machine (BRA) A voice bit saving bit total (B s) and precedence image use bit (B u) And it consists of means to calculate the bit total which can be used in order to encode the present image according to whenever [FIFO-buffer sufficiency] (B f). Then B s And B f A value is a bit rate control system according to claim 1 characterized by what is sent to said bit rate quota machine (BRA) when a demand signal is inputted into a voice coder and a FIFO buffer.

[Claim 5] Said Variable Bit Rate video encoder is a bit rate control system according to claim 1 characterized by having a means to control an image rate dynamically according to the bit which can be used for the present image coding, a means to control image coding mode dynamically, and the means that assigns a bit dynamically to said element according to the complexity (or activity) of the element of an image.

[Claim 6] The internal image coding mode in which the aforementioned dynamic image coding mode control means encodes an image only by the image component of itself (the image encoded by this approach here is I it is called an image), In the images coding mode (the image encoded by this approach here is P it is called an image) which encodes an image also from a precedence image using the image of itself to the predicting-coding approach The bit total B a which can be used, and min I The image coding bit rate B Imin and min P The image coding bit rate B Pmin is followed. I And P Bit rate control system according to claim 5 characterized by having a decision means to control image coding dynamically.

[Claim 7] The internal image coding mode in which the aforementioned dynamic image coding mode control means encodes an image only by the image component of itself (the image encoded by this approach here is I it is called an image), In the images coding mode (the image encoded by this approach here is P it is called an image) which encodes an image also from a precedence image using the image of itself to the predicting-coding approach The bit total B a which can be used, and P The presetting minimum bit total for image coding, And P excessive between the images and the present images which

were previously encoded according to the image modification component of the image inserted about the image encoded previously Bit rate control system according to claim 5 characterized by having a decision means to insert an image.

[Claim 8] said FIFO buffer -- whenever [buffer sufficiency] -- B f calculating -- and said BRA from -- bit rate control system according to claim 1 characterized by having a means to report the calculated value to BRA when a FIFO buffer receives a demand signal.

[Claim 9] Said multiplexer is a bit rate control system according to claim 1 characterized by having a means to generate the bit stream output by which the fixed bit rate was multiplexed by searching the bit stream of voice and video and adjusting the bit stream of a means to be mixed and to multiplex them to the synchronized bit stream, and voice and video, about those time amount and bit rate relation.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Industrial Application] This invention relates to the approach of assigning a video bit rate dynamically according to instant bit rate consumption of the sound signal especially in a fixed bit rate system about the video of a very low bit rate, and coding of related voice.

[0002]

[Description of the Prior Art] Digital-type video and coding of related voice play an important role in the application on the industry of storage of a digital signal, processing, transmission, and distribution. Various digital coding specification is enacted by various kinds of standardization organizations, or it is under establishment. In coding of an animation and related voice, the typical coding approach is 3. The part of **, i.e., video coding, voice coding, and system multiplex are included.

[0003] The present conversion coding algorithm adopted by the specification of the phase 1 of MPEG and a phase 2 includes technique, such as conversion, quantization, and variable length coding. In order to improve coding effectiveness, the predicting-coding technique, such as prediction between images and prediction of a motion, and amendment, is used. First, an image can be encoded by reducing the spatial redundancy in the image itself. Thus, the encoded image is usually I. It is called an image. An image can be encoded also in the mode between images again. It will be P if an image is predicted from a precedence image. It is called an image. If an image is predicted from both a precedence image and a consecutiveness image, it will point out a bidirectional predicting-coding image, and it is usually B. It is called an image. The main descriptions of the present video coding approach are to be fixed uniformly, when the bit rate of the encoded bit stream uses a FIFO buffer. Whenever [sufficiency / of a buffer] controls the bit rate of the each encoded image, and is I. An image and B An image and P It is used for adjusting the bit which can be used for an image, and a quantization step.

[0004] Many voice coding algorithms have appeared with a digital communication link technique, mobile communication technology, and the introductory technique of service of amusement or multimedia. Especially various very low bit rate voice coding algorithms are standardized in order to decrease the storage capacity of a transmission bit rate or a phonetic memory system. Long distance call quality voice is CCITT G.728 16-kilobit/standardized recently. Encoder of a second - It can obtain from a decoder. Communication link quality is USA. Federal Standard 1016 4.8-kilobit/It can obtain using the voice coder of a second. Vector Sum Excited Linear Predictive Coder standardized for [of North America and Japan] cell method communication links (vector sum exciting line form prediction encoder) 8.0 Kilobit/The thing near the long distance call quality in actuation of a second is offered. Generally these algorithms enter in the category of the voice coder known as a sign excitation encoder or a vector exciting line form prediction encoder, and have the intention of fixed bit rate transmission typically. The number of bits per voice frame is also maintained uniformly.

[0005] The Variable Bit Rate limited very much for voice coding is taken into consideration about the packet exchange network, the digital voice interpolation system, and the digital communication multiplexer system. The main functions of a system encoder are for the coded data in the buffer of a decoder to guarantee [to be in making it the need and offering sufficient information, in order to synchronize video, decode of speech information, and a display, and] overflow or not carrying out an

underflow to coincidence. Coding of system layer information includes packet-izing data to a packet, and creating the time stamp for packet headers. 2 The time stamp of ** is used. That is, it is the decode time stamp (DTS) which shows the time of day which decrypts the presentation time stamp (PTS) which shows the time of day which needs to perform or broadcast a voice frame or the presentation unit of a video image and voice, or a video image. PTS And DTS System clock criteria (SCR) It has the common time-of-day base called, and the measurement at the time of ** is unified, and it is made to ensure proper synchronization and buffer control. In the fixed fixed bit rate environment, a system needs the voice and video of a bit rate which were fixed for proper buffer control.

[0006]

[Problem(s) to be Solved by the Invention] However, the conventional coding approach mentioned above has the following trouble which bars achievement of the very high compression ratio which is an important factor for the coding application of the video of various very low bit rates, and related voice, and coding effectiveness.

[0007] That is, the non-efficiency of the present coding approach is produced from the following item.

1) The voice bit rate present fixed voice coding approach encodes a sound signal with a fixed bit rate. If we define speech information as the complexity measure which influences signal intensity and frequency distribution, complexity which is different with time amount will be found out. For example, if one person concerned with communicative is talking when a TV phone is considered, the person concerned with another side is usually hearing it silently. That is, there is the moment a hearer does not input voice. The 2nd A silent moment exists also in a message of people. A fixed voice coder wastes bandwidth at the moment of such silence. In the case of a very low bit rate coding application, voice may occupy bandwidth higher than it like video. The trouble which it is going to solve here is to use a Variable Bit Rate encoder for voice coding, and save the bit for voice coding.

2) The video bit rate present fixed video coding approach offers a fixed bit stream output by using a FIFO buffer for the end of an encoder. Whenever [instant bit use and buffer sufficiency] is used for adjusting the bit and quantization step for each images. The latter is used for controlling the bit rate of the macro block encoded by the degree in an image. This bit rate control process is carried out within the video encoder itself, and it is not related to bit use of a voice coder. How in order to improve video coding quality, bit saving from a voice coder is used is the trouble which it is going to solve.

3) The video dynamic bit rate control present video coding approach carries out bit rate control by assigning a fixed number of bits to each image, and adjusting a quantization step for every macro block of an image. There may be sufficient bit which can use more than which is sufficient for adjusting a quantization step to the minimum value. A bit excessive under such a condition cannot be used efficiently. other troubles -- the minimum quantization step -- the image in predetermined coding mode, i.e., I, An image and B An image and P a ***** [that it is the best approach of encoding an image] -- ** -- it is saying. The especially present image is B. The bit total which can be used when being set up so that it may encode by the image is P. An image or I It is B about a quantization step small even when it is sufficient amount to encode by the image. It is P to give an image. An image or I Does even an image give coded-image quality also with a good twist? An above-mentioned examination is I. An image and B An image and P It will come to a conclusion as a problem of how to select image coding mode dynamically.

[0008] The 2nd based on this item In the case of a very low bit rate with the image rate of a coding system lower than a rate required for the display of real-time video, a trouble is produced. In the case of a TV phone application, an image rate is usually set up by about 10 images per second as an example. At this image rate that decreased, if a motion of the body in an image is too quick, the development in which the motion was awkward will appear. This trouble makes predicting coding difficult like the time of a scene changing. How the bit which can be used in order to insert an image more than a predetermined image rate is used efficiently are other problems which should be taken up in order to improve image coding quality.

4) A system multiplex machine present system multiplex machine accepts only the voice of a fixed bit rate, and the bit stream of video, and multiplexes them to a fixed bit rate system bit stream. 2 The point which the device which controls the encoder of voice and video in assignment of the dynamic bandwidth between the encoders of ** does not have is a problem.

[0009] This invention is [preparing a communication link between the encoders of voice and video so that it may be efficiently used by the bit encoder, in order that bit saving obtained by voice coding may improve image quality in consideration of such a conventional technical problem, and] adjustable [I]. An image, B image, and P It aims at offering a much more flexible video bit rate controlling mechanism for the mode of an image, and dynamic image insertion.

[0010]

[Means for Solving the Problem] The dynamic bit rate video and the related voice coding system of this invention encode several frames of voice first. This voice frame number is decided with the image rate of a system. Subsequently, the system counts bit consumption and calculates bit saving from a voice coder during a precedence image coding period. The maximum number of bits can be saved during a silent period by supervising input voice level and analyzing the frequency and energy component. In the case of an owner sound period, the number of bits used for coding of each voice frame is selected so that the predetermined value or the necessary voice quality level of a signal pair noise with frequency weight may be guaranteed.

[0011] The number of bits saved with a voice coder is sent to a bit rate quota machine, and the quota machine investigates whenever [bit consumption / of the encoded precedence image /, and sufficiency / of the FIFO (FIFO) buffer], and calculates the bit which can be used for the present image coding based on these values.

[0012] The bit total which can be used for the present image is sent to the coding mode selection machine in a video encoder. This coding mode selection machine is I about the present image. An image and B An image or P It is determined to any of an image it encodes. Moreover, when [usually lower than an image rate] an image rate is needed by the real-time video application, it determines whether the excessive image should be inserted. This bit total that can be used is used also for controlling a quantization step by the present coding approach and the same approach, after image coding mode is selected.

[0013] A system multiplex machine makes it possible to compensate for the video delay to voice coding by inserting a time stamp in a system packet, in order to take the synchronization of the corresponding video and voice. Subsequently, the bit stream of the synchronized voice and video is multiplexed and forms a fixed bit rate system bit stream.

[0014]

[Function] At this invention, it is a bit saving counter (BSC 25), for example. Bit saving is calculated after each voice frame is encoded. Once BSC 25 receive a demand signal from the bit rate quota machine (BRA) 5, they will calculate the number of bits saved by the voice frame during the precedence image coding period. This saving bit total is shown as Bs. This bit saving total 4 It is a voice coder 2 for video coding. As a bit total which can be given, it is BRA 5. It is notified.

[0015] B u 7 of the bit total used for coincidence with the precedence image It is BRA 5 by the video encoder 10. It is supplied. It is B f 9 of whenever [sufficiency / of FIFO buffer 11] similarly. It is supplied by the detector whenever [in FIFO buffer 11 / sufficiency]. These 3 B s which is the parameter of **, and B u And B f Bit B a 8 which can be used subsequently to the present image coding Because of count, it is BRA 5. It is used.

[0016] BRA 5 B a 8 supplied According to a total, the video encoder 10 selects various image coding modes, and attains the best image coding quality considered. The output 12 of the encoded image is sent to FIFO buffer 11. The amount 7 of the bit used of the present image It is BRA 5 because of use of the next image coding. It is reported.

[0017] Encoded voice 3 Subsequently the bit stream of video 12 is multiplexed with the system multiplex vessel 14, and synchronization of voice and video is realized there. The system multiplex machine 14 multiplexes voice and video with a fixed bit rate, and generates the fixed system bit stream to a communication channel or a digital storage.

[0018]

[Example] Hereafter, the example of this invention is explained with reference to a drawing.

[0019] The coding system using the dynamic bit rate control of the combined voice and video coding in this invention is drawing. It is illustrated by 1. A coding system is a voice coder 2, the video encoder 10, the FIFO (FIFO) buffer 11, and the bit rate quota machine 5. And it has the system multiplex machine 14. This invention is a voice coder 2, the video encoder 10, and the bit rate quota machine 5, in order to support an operation of the whole coding system it not only to include such a system configuration and an operation, but. And the new means needed by the system multiplex machine 14 is included.

[0020] drawing 1 alike -- setting -- voice bit stream 3 Variable Bit Rate voice coder 1 The voice detection and discernment (ADI) to which it encodes and the encoder carries out some following tasks Module (Fig. 5) It has.

a) task b which detects the silent moment in a sound signal, and determines the voice frame skipped an owner sound segment -- mainly An owner voice segment and task c mainly classified into a non-voice segment Task d which assigns a bit based on the energy level with frequency weight of the analyzed message segment It is the bit saving counter (BSC) 25 (drawing 4) about bit consumption information.

The task to transmit, next a bit rate dynamic-control device are explained.

[0021] The desirable example of the coding system of this invention is drawing 1. It is shown, it sets there and a coding system is 2. Only the image coding mode of ** is included. The 1st Image coding mode is I. It is coding in an image which generates an image. The 2nd Image coding mode is P. It is predicting coding which generates an image. Prediction is only from a precedence image. However, the coding system of this invention is I. An image and P It is not limited only to the coding mode of an image. The system is I. An image and B An image and P It is applicable also to the coding mode of an image.

[0022] The controlling mechanism of dynamic bit rate assignment is drawing 2. It is shown. There is the maximum bit rate which presets for voice coding first, and it is shown as B_{amax}. Sound signal 1 The encoded voice bit rate is the constraint to which it is supposed that B_{amax} is not exceeded, and is the Variable Bit Rate encoder 2. It uses and encodes. A voice frame rate is F_a per second. It is set up and a video image rate is F_v per second. It is assumed that it is set up. A voice coder 2 is Period T_d. It is a sound signal 1 in between. Coding will be started. This time amount is delay needed by the video encoder 10, and is 1. It is set as an image period. The 1st It is BRA 5 just before starting coding of an image. It is B_a 8 to the video encoder 10. Delivery and the bit which can be used for the present image coding will be specified. The video encoder 10 is B_a 8, in order to apply suitable image coding mode. A value is analyzed.

[0023] The effectiveness which applies the above-mentioned rate control approach is drawing. It is shown also in 2. If the conventional coding approach which supplies a fixed bit rate video bit stream output is used, an average bit rate will be a value shown as B_{vave}. The first I An image is consecutive P. Quite many bits are consumed rather than an image (I image bit rate is 2 from P image bit rate usually twice is also high). By the coding approach of this example, it is the bit saving total B_s 4 of the voice frame in a precedence image period. It is used with the present image. Thereby, each image is a voice coder 2. Only a part to have been saved can be encoded in many bit totals too many. It will be drawing 2 if a silent period is in a sound signal. The whole bandwidth planned to voice can sometimes be used by video like [in the case of the 8th image which can be set]. These bits added are actually useful to a video encoder improving coded-image quality. average video bit rate B'_{vavg} increases at the beginning compared with the video bit rate B_{vavg} -- drawing 2 from -- it is clear. The following formula shows the relation between an average video bit rate and a voice bit rate.

[0024] B'_{vavg} - B_{vavg} = B_{amax} - B_{aavg}, however B_{aavg} are average voice bit rates.

[0025] Since a coding process is started in the delay period T_d to the start time of voice coding, and since a video bit rate is dynamically assigned again according to voice bit saving during the period of a precedence image, the video encoder 10 is the coding voice bit stream 3. Only the same period T_d shall be delayed so that it may be in agreement with the video bit stream 12 which corresponds in both a period and a bit rate. This operation is drawing 3. It is illustrated and bit saving in voice and the bit consumption in video are in agreement in time in there. This adjustment control is attained by the system multiplex machine 14, the bit stream of voice and video is alternatively taken out from FIFO buffer 11, and a time stamp proper for synchronization of voice and video is added.

[0026] drawing 2 ** -- drawing 3 from -- the remarkable description which is different from the conventional voice and the coding system of video so that it may understand -- voice bit stream 3 And it is that the video bit stream 12 is the thing of a Variable Bit Rate. However, voice bit stream 3 And the output system bit stream 15 after the video bit stream 12 was synchronized with the system multiplex vessel 14 is a fixed bit rate.

[0027] Next, a Variable Bit Rate voice coder is explained.

[0028] A bit saving counter and Variable Bit Rate voice coder 2 The relation of a between is drawing 4. It is shown concretely. Bit saving counter 25 (BSC) A sound signal 20 is F_a per second so that it may define above. It is segmented by the frame. i The bit consumption of the voice frame of eye watch is f_a. (i) It is 1 when it assumes that it is a bit 23. The bit saving total in the image period of ** can be calculated by following (several 1).

[0029]

[Equation 1]

N_{af}

$$B_s = B_{amax} \cdot N_{af} - \sum_{i=1} f_a(i)$$

i = 1

[0030] however, N_{af} 28 B_s a demand -- BAR 5 from -- before being made, it is the voice frame number

[0031] drawing 5 the purpose which asks for bit consumption for every voice frame so that it is alike and may be illustrated sake -- input message signal 101 Prediction analysis module 102 And audio detection and the module 103 of discernment It is passed. Prediction analysis module 102 Specified M The short-term linearity analysis of a spectrum which uses story autocorrelation analysis is carried out. Audio detection and module 103 of discernment The energy distribution passing through a speech frequency band is calculated by setting. When a non-sound is detected based on a threshold level, it is not necessary to carry out coding of a voice frame. prediction analysis module 102 from -- prediction coefficient 104 produced as a result a voice segment -- voiced or non-voice component (or still smaller classification) A classification is carried out. According to the class of application, the maximum bit rate which can be used for coding of each voice frame is the bit rate decision module 106 first. It can set and preset. It is equal to the maximum bit rate which can be used from a prediction coefficient and energy distribution information, or the individual bit rate level below it is based on other consciousness measure for coding of the measure of a signal pair noise with frequency weight, or a voice frame, and it is the bit rate decision module 106. It is set and selected. Bit consumption data 116 from each voice frame It is transmitted to a bit saving counter. Voice frame number 115 generated from a periodic counter It is transmitted to a frame counter. Coding with a bit rate lower than the maximum bit rate is a code book 108. A suitable subset is used (or the code book which decreased in number or the thing redesigned for [parameter] the methods of presentation). It can attain. Multiplex code book 108 The excitation generator which can generate various kinds of sequences including the sequence of the shape of the shape of a pulse and a noise is constituted. A total of 109 of an excitation sequence Prediction machine 110 which consisted of a spectrum prediction machine, various prediction machines, for example, pitch prediction machine, etc. It is passed. the predicted signal 111 -- input voice 101 from -- it deducts -- having -- difference 112, i.e., an error signal, It is generated and, subsequently weight is given in [this signal] consciousness. Error signal 114 by which the weight price was carried out in consciousness It is used for subsequently driving an error minimization vector quantization retrieval procedure. Rhine spectrum pair 118 And suitable parameter-sign 117 It is transmitted as a component of a voice bit stream.

[0033] Next, a bit rate quota machine (BRA) is explained. drawing 1 it is alike and is illustrated -- as -- BRA 5 Voice coder 2 from -- voice bit saving total B s 4 Video bit total B u 7 used with a precedence-prosign-ized image from the video encoder 10 whenever [and / sufficiency / of a FIFO buffer] -- B f 9 It receives. Image j It is B t about the presetting target number of bits of **. (j) It carries out and is the B t. (j) I Image coding and P Presetting I for each image coding The image coding bit rate B tI or presetting P It is j if it assumes that it is equal to the image coding bit rate B tP. The bit total which can be used for image coding of eye watch will be calculated by the following control means (several 2).

ID=000004

[0035] However, B_{fmax} is FIFO buffer size, and B_{AV} is a fixed system bit stream bit rate. j The number of bits which can be used for image coding of eye watch, (j-1) The target number of bits for the images of eye watch (j-1) To a difference with the actual number of bits used with the image of eye watch (j-1) j which applied the bit knot divisor of the voice coder in the coding period of the image of eye watch, and was given beforehand It is equal to what applied further the target number of bits for the images of eye watch. the number of bits B_a which can be used (j) Present FIFO whenever [buffer sufficiency] -- B_f (j) and the sum total of the maximum voice bit rate B_{amax} is larger than the maximum FIFO buffer size B_{fmax} -- if it becomes -- image j In order that the number of bits which can be used may avoid overflow of FIFO buffer 11 It will be calculated as what [below] deducted B_f (j) from the maximum FIFO buffer size

B fmax whenever [present FIFO-buffer sufficiency], and deducted the maximum voice bit rate B amax. the number of bits B a which can be used (j) Present FIFO whenever [buffer sufficiency] -- B f (j) and the sum total of the maximum voice bit rate B amax is smaller than the bit rate B AV of the combined voice and video -- if it becomes -- image j In order that the number of bits which can be used may avoid the underflow of FIFO buffer 11 From B AV to present FIFO It is B f whenever [buffer sufficiency]. (j) It will be calculated as what [more than] deducted and deducted the maximum voice bit rate B amax. [0036] Next, a Variable Bit Rate video encoder is explained.

[0037] The Variable Bit Rate video encoder 10 is drawing 6. It is constituted with the coding mode selection machine 31 and the bit consumption counter 35 so that it may be illustrated. The number of bits B a 30 which can be used for the present image is the coding mode selection machine (CMS) 31. It is sent. B Responding to the value of a 30, CMS 31 are I. Coding and P Coding or excessive P In order to select image coding, a decision-making process is carried out. This decision result is supplied to the Variable Bit Rate video encoder 33, in order to control coding of the present image 37. The encoded video bit stream is outputted through a link 34. an output 34 -- the bit consumption counter 35 -- being also alike -- since it is sent, the number of bits B u 36 used for the present image coding is calculated, and in order to use it for the next image coding, it is sent to BAR 5 (drawing 1).

[0038] The decision-making process of CMS 31 is drawing 7. It is shown by the flow chart. I They are B Imin and P about the minimum bit total required for coding of an image. The minimum bit total required for coding of an image is assumed to be B Pmin. Furthermore, threshold TH1 which it presets in order to inspect the image difference between the images located in the middle of the image with which precedence was encoded and the image with which precedence was encoded, and the present image encoded (PD) It is assumed that it is.

[0039] The bit total B a 50 which can be used for the present image is inputted into a comparator 51. B a 2 of B Pmin The bit total which it can use if larger than twice is 2. P of ** It appears in encoding an image enough and a certain thing is meant. In other words, it is excessive P. An image can be inserted. In this case, that comparator 54 is based on PD value defined previously by sending the output of a comparator 51 to a comparator 54 through a link 53, and it is P. It uses for inspecting further whether it is necessary to insert an image or not. PD value is the presetting threshold TH1. If large, it means that even the image located in the middle of the image with which precedence was encoded, and the present image encoded from the image with which precedence was encoded has a substantial change in an image component. therefore, the decision process 62 -- the output of a comparator 54 -- a link 58 -- a passage -- "Yes" -- if it becomes -- 2 P of ** It carries out so that an image may be encoded.

[0040] However, if a comparator 54 "No" Comes to pass a link 59, it will not have a substantial change about the image with which precedence was encoded, and it is excessive P. It means that it is not necessary to insert an image. Bit total B a which can be used Subsequently it is 1. I of ** Coding of an image, or 1 P of ** It is inspected through a link 59 by the comparator 55 for decision making of coding of an image. This process is applied also when a comparator 51 outputs "No" from a link 52. Therefore, the continuing decision-making procedure is applied to the case of a link 52 and a link 59.

[0041] Bit total B a which can be used in a comparator 55 Min I It is compared with the image coding bit total B Imin. B a >= B Imin The bit total which it can use if it becomes is 1. I of ** It is enough to encode an image and is I of the decision process 60. It means that image coding can be started from a link 56. B a -- -- B Imin if it becomes -- I of the decision process 61 Image coding is selected by the output from a link 57.

[0042] 1 of the decision result of 60, 61, and 62 Only ** is selected at once and it is 63, 64, and 1 of the link of 65. It is sent to the coding mode encoder 66 through **, and the encoder outputs the selected decision result to an output link 67.

[0043] The device mentioned above is dynamic I. An image and P It is defined as image coding modal control.

[0044] It is B a if image coding mode is decided. It is used for adjusting a quantization step so that the bit rate generated from the present image coding may be controlled dynamically. This device is defined as dynamic quantization step control.

[0045] Subsequently, the Variable Bit Rate video encoder 33 is used for encoding the present image according to the image coding mode selected by CMS 31. 1 of the video encoder 33 The desirable example of ** is drawing 8. It is the conversion coding system shown.

[0046] The input video signal containing the pixel data of each image is inputted and stored in the input frame storage 71. The block sampling circuit 72 receives the data stored in the input frame store 71 through Rhine 84, and divides the image data into a non-overlapping block spatially [pixel data]. In

order to acquire the adaptability of appropriate level, it is $\lceil 8 \rceil$. The block size of a pixel is used. The frame number of an image is inputted also into the input image store 71, and is passed to a switch (S1) 73 through the block sampling circuit 72. Based on a frame number, a switch 73 lets Rhine 86 pass for coding with the coding mode in an image, or sends out the output block of pixel data from the block sampling circuit 72 alternatively to Rhine 87 for coding with predicting-coding mode.

[0047] In the case of the coding mode in an image, the output of the block sampling circuit 72 is passed to the discrete cosine transformation (DCT) circuit 76 through Rhine 86. DCT A circuit 76 carries out discrete cosine transformation which is mathematical conversion of the common knowledge which changes image data into the data in a frequency domain. The changed data, i.e., DCT, A multiplier receives the process of quantization in the quantization circuit 77 which uses the quantization step size subsequently given through Rhine 93 by the quantization matrix and the rate control unit 79. As for the quantized data, it passes along Rhine 91 with the quantization step as auxiliary information, and run length coding and the variable-length coding network 78 are passed, and the circuit carries out run length coding of the quantized data and its related auxiliary information, and, subsequently carries out variable length coding of a run-length-coding result. encoded bit stream 100 which the output of run length coding and the variable-length coding network 78 can transmit to a decoder it is. This bit stream is passed to the rate control unit 79 through Rhine 92 again. Based on the number of bits already used at the time of coding of a block, the rate control device 79 adjusts a quantization step so that an output bit stream may satisfy the bit rate requirement of a coding system. The value quantized by the quantization circuit 77 lets Rhine 94 pass, and is passed to the reverse quantization circuit 80 and the reverse discrete cosine transformation (reverse DCT) circuit 81 again. The reverse quantization circuit 80 and reverse DCT A circuit is DCT. Although a circuit 76 and the quantization circuit 77 carry out, a reverse process is carried out and the reconfigured data are obtained. The reconfigured data are stored in the partial decryption image storage 82 through Rhine 96, and are outputted to the prediction circuit and amendment circuit of the motion for the following input image which is a predicting-coding image. I After the whole image was encoded, a rate control device is I. The sum total bit used for coding of an image is calculated, and it is the following P. An image and I The number of bits assigned to an image is determined. the case in predicting-coding mode -- the output of the block sampling circuit 72 -- a switch S1 -- and subsequently it lets Rhine 87 pass and the motion vector (MV) prediction circuit 74 is passed. In MV prediction circuit 74, motion vector prediction uses the partial decryption image from the partial decryption image storage 82 obtained through Rhine 98, and it is carried out so that MV may be decided from a contiguity image. MV prediction circuit 74 finds out the best coincidence by finding out the block which was in agreement with best based on some predetermined standards like for example, a mean square error, and searching for the shift motion direction of a block of the pixel from one image to the following image.

[0048] MV obtained in the motion vector prediction process is passed to the motion amendment circuit 75 through Rhine 88 with the block encoded, and the circuit amends by moving using the block stored in the partial decryption image store 82 in order to search for the predicted block. A prediction block is applied to the block which can acquire the offset given by MV from a partial decryption store, and is acquired. the difference between the value of the pixel of the block encoded and the value of the pixel of a prediction block calculates -- having -- difference -- the difference which consists of a value -- a block is acquired. difference -- the standard which defined the motion amendment circuit 75 as some points, such as a standard deviation of a block, again based on the block -- using it -- difference -- comparing the complexity (activity) of a block and the block encoded -- a conversion sake -- DCT It judges whether it is necessary to move and amend the block passed to a circuit 76. if the block changed is judged that it is necessary to move and amend -- the difference -- a block -- Rhine 89 -- letting it pass -- DCT A circuit 76 is passed. If the block changed is judged that it is not necessary to move and amend, it will be the block (namely, the original block outputted from the block sampling circuit 72) itself. It lets Rhine 89 pass and is DCT. A circuit 76 is passed (it is the same as that of the block of the coded image in an image). The information bit which shows whether the block moved the amendment circuit 75 and it encoded by amendment is generated by furthermore moving, and it is DCT in a motion vector about an information bit. A circuit 76 is passed. namely, Rhine 89 -- letting it pass -- moving -- the amendment circuit 75 to DCT the data outputted to a circuit 76 -- difference -- block (motion amendment was carried out and it encoded) Or the original block (it encoded without motion amendment) it is -- the information on whether motion amendment of the information bit was carried out, and no, and difference -- in a block, the information on a motion vector is shown.

[0049] The block data from the motion amendment circuit 75 is DCT. It sets in a circuit 76 and is DCT. It is changed into a multiplier and is DCT. It quantizes in the quantization circuit 77 and a multiplier is

Quantization DCT. It becomes a multiplier. The information bit which shows, respectively whether the block moved and it encoded by amendment or MV is DCT. A circuit 76 is passed through a passage and Rhine 90 in the quantization circuit 77. the quantized data (quantized DCT multiplier) run length coding and the variable-length coding network 78 are passed with the information bit which shows, respectively whether the auxiliary information containing a quantization step, a motion vector, and motion amendment were carried out to the block -- having -- run length coding -- and variable length coding is carried out and it becomes an output coding bit stream. The bit stream is passed to the rate control unit 79 through Rhine 92 again. It is passed through Rhine 94, and reverse-quantizes in the reverse quantization circuit 80 again, and the quantization data from the quantization circuit 77 are reverse DCT. It sets in a circuit 81 and is reverse DCT. It is carried out and is stored in the partial coded-image storage 82 for coding of the following image.

[0050] Next, a FIFO buffer is explained.

[0051] Drawing 9 An example of the FIFO-buffer condition in the coding process of **, voice, and video is shown. The 1st Period T d Between and coding voice bit stream 3 It is written in FIFO buffer 11. moment T d from -- starting -- FIFO buffer 11 -- voice bit stream 3 And the video bit stream 12 is sufficient. FIFO buffer 11 -- whenever [presetting sufficiency] -- B O up to -- if sufficient, the system multiplex machine 14 will start collecting bits from FIFO buffer 11. Since the system multiplex machine 14 takes out a bit from FIFO buffer 11 with the fixed bit rate B AV, whenever [buffer sufficiency] is B out. The fixed total which is shown by carrying out decreases. This value is always fixed. Since the bit rate of voice and video is adjustable, those combination is also Variable Bit Rates. Therefore, the rate of FIFO buffer 11 is drawing 9 by the inclination where curves differ whenever [in each period / buffer sufficiency]. It is 1 so that it may be illustrated. It is changing to other periods from the period of **. B out It is set up so that overflow or the underflow of FIFO buffer 11 may be avoided. However, this is also drawing 9. Time delay is produced in the bit stream transmitted so that it may be shown by 3T d.

[0052] Next, a system multiplex machine is explained.

[0053] A system multiplex machine multiplexes the video and voice data of the input from a FIFO buffer by inserting a time stamp in a voice frame and a video image, in order to ensure synchronization proper [depending especially] which packet-izes the video and voice data of an input to the packet of fixed die length. 2 The time stamp of ** is used. namely, decode time stamp (DTS) which shows the time of day which decrypts the presentation time stamp (PTS) which shows the time of day which needs to perform or broadcast a voice frame or the presentation unit of a video image and voice, or a video image it is . PTS And DTS System clock criteria (SCR) It has the common time-of-day base called, and the measurement at the time of ** is unified, and it is made to ensure proper synchronization and buffer control.

[0054] It is suitable PTS, if a trace of the number of images to which a system multiplex machine is sent in the case of voice data is continued and new data are obtained from a FIFO buffer. It is inserted in a packet header. DTS Since display time of day has decryption time of day and fixed relation, it is not needed.

[0055] In the case of a video data, a system multiplex machine continues a trace of the number of images sent, and is suitable PTS. And DTS It is inserted in a packet header. DTS It is required when the decryption time of day and display time of day of an image differ from each other.

[0056]

[Effect of the Invention] This invention can improve the quality of a video signal sharply, without reaching a compromise about voice quality, since the bit which can be used with image coding mode through the period fixed by offering the device which controls an image rate can be used much more efficiently so that clearly from the place described above.

[0057] This invention is effective for the video of a very low bit rate, and especially coding of related voice. For example, when this invention is applied to a TV phone application, by using all the voice bit rates of those who have not talked, the man's image is encoded in more bits, it comes to be transmitted to the partner under message, and image quality can be improved.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] The block diagram of the video by which bit rate control is carried out dynamically, and the

coding system of related voice is shown.

[Drawing 2] Voice, the time amount in the coding process of video, and the relation of a bit rate are illustrated.

[Drawing 3] The time amount in a multiplexing process and the relation of a bit rate are illustrated.

[Drawing 4] The relation between a bit saving counter and a Variable Bit Rate voice coder is shown.

[Drawing 5] The block diagram of a Variable Bit Rate voice coder is shown.

[Drawing 6] With a coding mode selection machine and a bit consumption counter, the block diagram of the configuration of a Variable Bit Rate video encoder is shown.

[Drawing 7] The device of a dynamic image coding mode selection flow chart is illustrated.

[Drawing 8] The block diagram of the video conversion coding approach is shown.

[Drawing 9] Whenever [FIFO-buffer sufficiency] is illustrated.

[Description of Notations]

1 Sound Signal

2 Voice Coder

3 Voice Bit Stream

5 Bit Rate Allocation Machine

6 Video Signal

10 Video Encoder

11 FIFO Buffer

12 Video Bit Stream

14 System Multiplex Machine

15 System Bit Stream

B u : Bit total used by the precedence video frame

B s : Bit total saved by the voice frame in front of the present video frame

B f : Whenever [before the present video frame / FIFO-buffer sufficiency]

B a : The number of bits which can be used for the present frame

B AV : Fixed bit rate which the bit stream of the combined voice and video presets

B amax: The maximum voice bit rate

B aavg: Average voice bit rate

B vavg: Average video bit rate

B'vavg: New average video bit rate after adding the bit saved by the voice coder

T d : Video time delay over the start point of a voice coder

** The number of bits from a voice coder

[Translation done.]

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平7-202820

(43)公開日 平成7年(1995)8月4日

(51)Int.Cl.⁸

H 0 4 B 14/04

H 0 4 N 7/30

9/802

識別記号

Z 9372-5K

片内整理番号

F I

技術表示箇所

H 0 4 N 7/ 133

9/ 80

審査請求 未請求 請求項の数 9 O L (全 12 頁)

Z

F

(21)出願番号

特願平5-336975

(22)出願日

平成5年(1993)12月28日

(71)出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72)発明者

ファン・シー ユン

シンガポール 1646 #13-404 ベドッ

クノース ストリート 3 ブロック

525

(72)発明者

タン・オー ベン

シンガポール 1955 #10-03 セラング

ーノース アベニュー 4 ブロック

542

(74)代理人

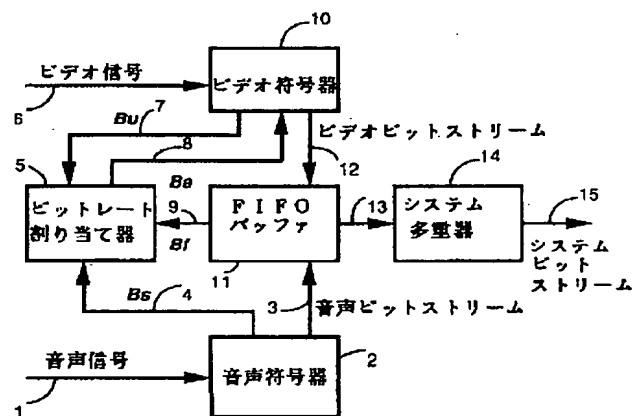
弁理士 松田 正道

(54)【発明の名称】 ビットレート制御システム

(57)【要約】

【目的】 音声符号化により得られるビット節約が、画像品質を向上するためにビット符号器により効率的に使用されるように、音声とビデオの符号器間に通信リンクを設けるシステムを提供すること。

【構成】 動的にビットレートを制御するビットレート制御システムであって、音声信号を符号化し、ビデオ符号器用に使用し得る、音声符号器で節約されたビット総数を与える可変ビットレート音声符号器2と、毎秒一定数の画像単位で入力ビデオフレームを符号化する可変ビットレートビデオ符号器10と、可変ビットレート音声符号器の出力を利用して、符号化される現行画像に利用できるビット総数を動的に割り当てるビットレート割り当て器5と、符号化された音声とビデオのビットストリームをバッファリングし、組合わせられたビデオおよび関連する音声の一定ビットレートビットストリームを与えるFIFOバッファ11と、音声とビデオの信号を同期化するために加えられる時刻表示で、音声とビデオのビットストリームを多重化する多重装置14とを備える。



Bu: 先行ビデオフレームにより使用されたビット総数

Bs: 現行ビデオフレームの前に音声フレームにより節約されたビット総数

Ba: 現行ビデオフレームの前のFIFOバッファ充足度

Ba: 現行フレームに利用できるビット数

【特許請求の範囲】

【請求項 1】 動的にビットレートを制御するビットレート制御システムであって、
音声信号を符号化し、ビデオ符号器用の節約ビット総数 (B_s) を与える可変ビットレート音声符号器と、
毎秒一定数の画像単位で入力ビデオシーケンス信号を付号化する可変ビットレートビデオ符号器と、
前記可変ビットレート音声符号器の出力を利用して、符号化される現行画像に利用できるビット総数を動的に割り当てるビットレート割り当て器と、

符号化された音声とビデオのビットストリームをバッファリングし、組合わせられたビデオおよび関連する音声の一定ビットレートビットストリームを与えるFIFOバッファと、

音声とビデオの信号を同期化するために加えられる時刻表示で、音声とビデオのビットストリームを多重化する多重装置と、

を備えたことを特徴とするビットレート制御システム。

【請求項 2】 前記可変ビットレート音声符号器は、
音声予測分析と連係して、音声フレームが、音声フレームの符号化に使用する必要がある適切なビットレートを必要とするか、またはスキップされ得るかを事前判断する、音声検出・識別手段 (ADI) と、
先行画像符号化期間中に、節約されたビット総数 (B_s) を計算し、ビットレート割り当て器 (BRA) へそれを送る、合計音声ビットレート計算手段と、
を備えたことを特徴とする請求項 1 記載のビットレート制御システム。

【請求項 3】 前記音声検出・識別手段 (ADI) は、
エネルギーレベルとエネルギーの分布を求めるために、
入力音声信号の周波数およびエネルギー成分を分析する手段と、
前記のエネルギーレベルとエネルギーの分布から音声信号種類を、無音期間か、主として無声音信号か、または主として声音信号に分類する手段と、
音声符号化のために所定のビットレートを生じるように、許容される符号化不正確さ、および量子化ノイズを予測する手段と、
を備えたことを特徴とする請求項 2 記載のビットレート制御システム。

【請求項 4】 前記ビットレート割り当て器 (BRA) は、
音声ビット節約ビット総数 (B_s)、先行画像使用ビット (B_u) およびFIFOバッファ充足度 (B_f) に従って現行画像を符号化するために利用できるビット総数を計算する手段から構成され、
それら B_s および B_f 値は、要求信号が音声符号器およびFIFOバッファへ入力されるときに、前記ビットレート割り当て器 (BRA) へ送られる、
ことを特徴とする請求項 1 記載のビットレート制御システム。

【請求項 5】 前記可変ビットレートビデオ符号器は、
現行画像符号化に利用できるビットに従って画像レートを動的に制御する手段と、
画像符号化モードを動的に制御する手段と、
画像の小部分の複雑さ (又は活性度) に従って前記小部分にビットを動的に割り当てる手段と、
を備えたことを特徴とする請求項 1 記載のビットレート制御システム。

【請求項 6】 前記の動的な画像符号化モード制御手段は、

それ自体の画像成分だけにより画像を符号化する内部画像符号化モード (ここに、この方法により符号化される画像はI 画像と呼ぶ) と、それ自体の画像からだけではなく、予測符号化方法を使用して先行画像からも画像を符号化する画像間符号化モード (ここに、この方法により符号化される画像はP 画像と呼ぶ) とにおいて、
利用できるビット総数 B_a 、最小I 画像符号化ビットレート B_{Imin} および最小P 画像符号化ビットレート B_{Pmin} に従って、I およびP の画像符号化を動的に制御する決定手段を備えたことを特徴とする請求項 5 記載のビットレート制御システム。

【請求項 7】 前記の動的な画像符号化モード制御手段は、

それ自体の画像成分だけにより画像を符号化する内部画像符号化モード (ここに、この方法により符号化される画像はI 画像と呼ぶ) と、それ自体の画像からだけではなく、予測符号化方法を使用して先行画像からも画像を符号化する画像間符号化モード (ここに、この方法により符号化される画像はP 画像と呼ぶ) とにおいて、
利用できるビット総数 B_a 、P 画像符号化のための事前設定最小ビット総数、および先に符号化された画像に関して挿入される画像の画像変更成分に従って、先に符号化された画像と現行画像との間に余分のP 画像を挿入する決定手段を備えたことを特徴とする請求項 5 記載のビットレート制御システム。

【請求項 8】 前記FIFOバッファは、
バッファ充足度 B_f を計算し、かつ前記BRA からの要求信号をFIFOバッファが受信するときに、その計算値をBRAへ報告する手段を備えたことを特徴とする請求項 1 記載のビットレート制御システム。

【請求項 9】 前記多重装置は、
音声とビデオのビットストリームを検索し、かつそれらを混合し、同期化されたビットストリームに多重化する手段と、
音声とビデオのビットストリームを、それらの時間とビットレート関係に関して整合させることにより、一定ビットレートの多重化されたビットストリーム出力を発生する手段とを備えたことを特徴とする請求項 1 記載のビットレート制御システム。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、非常に低いビットレートのビデオと関連音声の符号化に関し、特に一定ビットレートシステムにおける音声信号の瞬時ビットレート消費に従ってビデオビットレートを動的に割り当てる方法に関するものである。

【0002】

【従来の技術】ディジタル式のビデオと関連音声の符号化は、ディジタル信号の記憶、処理、伝送および分配という産業上の用途において重要な役割を果たす。各種の標準化組織により種々のディジタル符号化規格が制定され、又は制定中である。動画と関連音声の符号化の場合、典型的な符号化方法は、3つの部分、すなわちビデオ符号化、音声符号化およびシステム多重化を包含する。

【0003】MPEGのフェーズ1とフェーズ2の規格により採用される現行の変換符号化アルゴリズムは、変換、量子化および可変長符号化などの手法を包含する。符号化効率を向上するために、画像間予測、および動きの予測と補正などの予測符号化手法が使用される。まず、画像は、画像自体内の空間的冗長さを減らすことにより符号化できる。このようにして符号化された画像は、通常I画像と呼ばれる。画像は又、画像間モードでも符号化できる。画像が先行画像から予測されるならば、それはP画像と呼ばれる。画像が先行画像と後続画像との両方から予測されるならば、それは双方向予測符号化画像を指し、通常B画像と呼ばれる。現行のビデオ符号化方法の主な特徴は、符号化されたビットストリームのビットレートがFIFOバッファを使用することにより一定に固定されるということにある。バッファの充足度は、各符号化された画像のビットレートを制御して、I画像、B画像およびP画像に利用できるビットと量子化ステップを調整するのに使用される。

【0004】多くの音声符号化アルゴリズムは、ディジタル通信リンク技術、移動体通信技術、娯楽やマルチメディアのサービスの導入技術と共に出現している。特に種々の非常に低いビットレート音声符号化アルゴリズムは、伝送ビットレート、または音声記憶システムの記憶容量を減少するために標準化されている。市外通話品質音声は、最近標準化されたCCITT G.728 16キロビット/秒の符号器-復号器から得ることができる。通信品質は、USA 連邦規格1016 4.8キロビット/秒の音声符号器を使用して得ることができる。北米と日本のセル方式通信向けに標準化されたVector Sum Excited Linear Predictive Coder(ベクトル和励起線形予測符号器)は、8.0キロビット/秒の作動における市外通話品質に近いものを提供する。これらのアルゴリズムは一般的に、符号励起符号器またはベクトル励起線形予測符号器として知られる音声符号器の部類に一般的に入り、また典型的には一定ビットレート伝送を意図している。音声フレーム

当たりのビット数も一定に維持される。

【0005】音声符号化のために非常に限定された可変ビットレートは、パケット交換網、ディジタル音声補間システムおよびディジタル通信多重化装置システムについて考慮されている。システム符号器の主な機能は、ビデオと音声情報の復号と表示を同期化するために必要にして十分な情報を提供することであり、また同時に復号器のバッファにおける符号化データがオーバーフローまたはアンダーフローしないことを保証することにある。システムレイヤ情報の符号化は、データをパケットにパケット化すること、およびパケットヘッダー用の時刻表示を作成することを含む。2つの時刻表示が使用される。すなわち、音声フレームまたはビデオ画像の提示ユニットを演奏または放送する必要がある時刻を示す提示時刻表示(PTS)、および音声またはビデオ画像を復号化する時刻を示す復号時刻表示(DTS)である。PTSおよびDTSは、システムクロック基準(SCR)と呼ばれる共通時刻ベースを有して、測時の計測を統一し、かつ適正な同期化およびバッファ管理を確実にするようにする。固定された一定ビットレート環境においては、システムは、適正なバッファ管理のために固定されたビットレートの音声とビデオを必要とする。

【0006】

【発明が解決しようとする課題】しかし、上述した従来の符号化方法は、種々の非常に低いビットレートのビデオと関連音声の符号化用途にとり、重要な要因である非常に高い圧縮比および符号化効率の達成を妨げる下記の問題点を有する。

【0007】すなわち、現行の符号化方法の非効率率は下記項目から生じる。

1) 一定の音声ビットレート

現行の音声符号化方法は、一定ビットレートで音声信号を符号化する。我々が音声情報を、信号の強さと周波数分布に影響する複雑性測度と定義するならば、時間と共に異なる複雑性を見出す。例えばテレビ電話を考えた場合、通信の一方の当事者が話していると、他方の当事者は通常無言で聞いている。すなわち聞き手が音声を入力しない瞬間がある。第2に、人の通話の中にも無言の瞬間が存在する。一定音声符号器は、これらの無言の瞬間に帯域幅を浪費する。非常に低いビットレート符号化用途の場合、音声はビデオと同様、またはそれより高い帯域幅を占めることがある。ここで解決しようとする問題点は、音声符号化に可変ビットレート符号器を使用し、かつ音声符号化用のビットを節約することにある。

2) 一定のビデオビットレート

現行のビデオ符号化方法は、符号器の末端部にFIFOバッファを使用することにより一定ビットストリーム出力を提供する。瞬時ビット使用およびバッファ充足度は、各画像用のビットと量子化ステップを調整するのに使用される。後者は、画像内の次に符号化されるマクロブロッ

クのビットレートを制御するのに使用される。このビットレート制御プロセスは、ビデオ符号器自体内で実施され、それは音声符号器のビット使用とは関係ない。ビデオ符号化品質を向上するために音声符号器からのビット節約をどのようにして利用するかが、解決しようとする問題点である。

3) ビデオ動的ビットレート制御

現行のビデオ符号化方法は、一定数のビットを各画像に割り当て、かつ画像のマクロブロック毎に量子化ステップを調整することによりビットレート制御を実施する。量子化ステップを、その最小値まで調整するに足る以上の利用できる十分なビットがある場合がある。そのような状態下で、余分のビットは効率的には使用できない。他の問題点は、最小の量子化ステップが、所定の符号化モードでの画像、すなわちI 画像、B 画像およびP 画像を符号化する最良の方法であるかどうかということである。特に現行の画像がB 画像で符号化されるように設定されているときに、利用できるビット総数がP 画像またはI 画像で符号化するに十分な量である場合でも、小さい量子化ステップをB 画像に与えることが、P 画像よりも、またはI 画像さえよりも良好な符号化画像品質を与えるであろうか。上述の検討は、I 画像、B 画像およびP 画像符号化モードをどのようにして動的に選定するかの問題として結論付けられるであろう。

【0008】この項目に基づく第2の問題点は、符号化システムの画像レートが実時間ビデオの表示に必要なレートよりも低い、非常に低いビットレートの場合に生じる。一例としてテレビ電話用途の場合に画像レートは、毎秒ほぼ10画像に通常設定される。この減少した画像レートでは、画像中の物体の動きが速過ぎるならば、動きがギクシャクした現像が現れるであろう。この問題点は、場面が変わるときと同様に、予測符号化を困難にする。画像を所定の画像レート以上に挿入するために利用できるビットを如何に効率的に利用するかは、画像符号化品質を向上するために取り上げるべき他の問題である。

4) システム多重器

現行のシステム多重器は、一定ビットレートの音声とビデオのビットストリームだけを受入れて、それらを一定ビットレートシステムビットストリームに多重化する。2つの符号器間の動的帯域幅の割り当ての場合に音声とビデオの符号器を制御する機構はない点が問題である。

【0009】本発明は、このような従来の課題を考慮し、音声符号化により得られるビット節約が、画像品質を向上するためにビット符号器により効率的に使用されるように、音声とビデオの符号器間に通信リンクを設けること、また、可変I 画像、B画像、P 画像のモードおよび動的画像挿入のため、一層柔軟なビデオビットレート制御機構を提供することを目的とするものである。

【0010】

【課題を解決するための手段】本発明の、動的ビットレートビデオと関連音声符号化システムは、先ず音声の数フレームを符号化する。この音声フレーム数は、システムの画像レートにより決められる。ついで、そのシステムは、ビット消費をカウントし、また先行画像符号化期間中に音声符号器からのビット節約を計算する。入力音声レベルを監視し、かつその周波数とエネルギー成分を分析することにより、無音期間中に最大のビット数が節約できる。有音期間の場合、各音声フレームの符号化に使用されるビット数は、所定の周波数重みつき信号対ノイズの値、または所要の音声品質レベルを保証するように選定される。

【0011】音声符号器で節約されるビット数は、ビットレート割り当て器へ送られ、その割り当て器は、符号化された先行画像のビット消費量および先入れ先出し(FIFO)バッファの充足度を調べ、これらの値をもとに現行画像符号化に利用できるビットを計算する。

【0012】現行画像に利用できるビット総数は、ビデオ符号器内の符号化モード選定器へ送られる。この符号化モード選定器は、現行画像をI 画像、B 画像又はP 画像のいずれに符号化するかを決定する。又、画像レートが実時間ビデオ用途により必要とされる通常画像レートよりも低いときに余分の画像を挿入すべきであるか否かも決定する。この利用できるビット総数は、画像符号化モードが選定された後、現行符号化方法と同様な方法で量子化ステップを制御するのにも使用される。

【0013】システム多重器は、該当するビデオと音声の同期をとる為に時刻表示をシステムパケットに挿入することにより、音声符号化に対するビデオ遅延を埋め合わせることを可能にする。ついで同期化された音声とビデオのビットストリームは多重化されて、一定ビットレートシステムビットストリームを形成する。

【0014】

【作用】本発明では、例えば、ビット節約計数器(BSC 25)は、各音声フレームが符号化された後にビット節約を計算する。BSC 25は、ビットレート割り当て器(BRA)5から要求信号を一旦受信すると、先行画像符号化期間中に音声フレームにより節約されたビット数を計算する。この節約ビット総数は B_s として示される。このビット節約総数4は、ビデオ符号化用に音声符号器2により与え得るビット総数として、BRA 5に通知される。

【0015】同時に、先行画像により使用されたビット総数の B_u 7は、ビデオ符号器10によりBRA 5に供給される。同様にFIFOバッファ11の充足度の B_f 9も、FIFOバッファ11内の充足度検出器により供給される。これらの3つのパラメータである B_s 、 B_u および B_f はついで、現行画像符号化に利用できるビット B_a 8の計算のために、BRA 5により使用される。

【0016】BRA 5により供給される B_a 8の総数に従って、ビデオ符号器10は種々の画像符号化モードを選定

して、考えられる最良の画像符号化品質を達成するようにする。符号化された画像の出力12は、FIFOバッファ11へ送られる。現行画像のビット使用量7は、次の画像符号化の使用のためにBRA 5に報告される。

【0017】符号化された音声3とビデオ12のビットストリームはついで、システム多重器14により多重化され、そこで音声とビデオの同期化が実現される。システム多重器14は、一定ビットレートで音声とビデオを多重化して、通信チャネルまたはデジタル記憶媒体への一定システムビットストリームを生成する。

【0018】

【実施例】以下、本発明の実施例について図面を参照して説明する。

【0019】本発明における、組合わせた音声とビデオ符号化の動的ビットレート制御を用いた符号化システムは、例えば、図1に図示される。符号化システムは、音声符号器2、ビデオ符号器10、先入れ先出し(FIFO)バッファ11、ビットレート割り当て器5およびシステム多重器14を備える。本発明は、そのようなシステム構成と作用を包含するだけでなく、全体の符号化システムの作用をサポートするために、音声符号器2、ビデオ符号器10、ビットレート割り当て器5およびシステム多重器14により必要とされる新規の手段を包含する。

【0020】図1において、音声ビットストリーム3は、可変ビットレート音声符号器1により符号化され、その符号器は、下記の幾つかのタスクを実施する音声検出と識別(ADI)モジュール(図5)を備える。

- a) 音声信号における無音の瞬間を検出し、かつスキップされる音声フレームを決定するタスク
- b) 有音セグメントを、主として有音声セグメントと主として無音声セグメントに分類するタスク
- c) 分析された通話セグメントの周波数重みつきエネルギーレベルに基づいてビットを割り当てるタスク
- d) ビット消費情報をビット節約計数器(BSC)25(図4)へ転送するタスク

次に、ビットレート動的制御機構について説明する。

【0021】本発明の符号化システムの好ましい実施例は図1に示され、そこにおいて符号化システムは、2つの画像符号化モードだけを包含する。第1の画像符号化モードは、I画像を発生する画像内符号化である。第2の画像符号化モードは、P画像を発生する予測符号化である。予測は先行画像からだけである。しかしながら本発明の符号化システムは、I画像およびP画像の符号化モードだけに限定されない。そのシステムは、I画像、B画像およびP画像の符号化モードにも適用できる。

【0022】動的ビットレート割り当ての制御機構は図2に示される。まず音声符号化用の事前設定される最大ビットレートがあり、 B_{amax} として示される。音声信号1は、符号化された音声ビットレートは B_{amax} を越えないとする制約で、可変ビットレート符号器2を使用して

符号化される。音声フレームレートは毎秒 F_a に設定され、またビデオ画像レートは毎秒 F_v に設定されると仮定する。音声符号器2は、期間 T_d の間に音声信号1の符号化を開始することになる。この時間は、ビデオ符号器10により必要とされる遅延であり、1画像期間に設定される。第1の画像の符号化を開始する直前に、BRA 5はビデオ符号器10に $B_a 8$ を送り、現行画像符号化に利用できるビットを指定することになる。ビデオ符号器10は、適切な画像符号化モードを適用する為に、 $B_a 8$ の値を分析する。

【0023】上記のレート制御方法を適用する効果は、図2にも示される。一定のビットレートビデオビットストリーム出力を供給する従来の符号化方法が使用されるならば、平均ビットレートは、 B_{vave} として示される値であろう。最初のI画像は、後続のP画像よりもかなり多くのビットを消費する(I画像ビットレートはP画像ビットレートよりも2倍も高いのが通常である)。本実施例の符号化方法で、先行画像期間中の音声フレームのビット節約総数 $B_s 4$ は現行画像により使用される。これにより、各画像は、音声符号器2により節約された分だけ余計に多くのビット総数で符号化できる。音声信号に無音期間があるならば、図2における8番目の画像の場合のように、音声に予定した全体帯域幅をビデオにより使用できることが時々ある。これらの付加されるビットは、ビデオ符号器が符号化画像品質を向上するのに実際に役立つ。平均ビデオビットレート B'_{vavg} は、当初ビデオビットレート B_{vavg} と比べて増加されることは図2から明らかである。下記の式は、平均ビデオビットレートと音声ビットレートとの関係を示す。

【0024】 $B'_{vavg} - B_{vavg} = B_{amax} - B_{aavg}$
ただし B_{aavg} は平均音声ビットレートである。

【0025】ビデオ符号器10は、音声符号化の開始時間に対して遅延期間 T_d で符号化プロセスを開始するので、かつまたビデオビットレートは、先行画像の期間中の音声ビット節約に従って動的に割り当てられるので、符号化音声ビットストリーム3は、期間とビットレートの両方において対応するビデオビットストリーム12に一致するように、同一の期間 T_d だけ遅延されるものとする。この演算は図3に図示され、そこにおいて音声におけるビット節約とビデオにおけるビット消費は時間的に一致している。この整合制御はシステム多重器14により達成され、FIFOバッファ11から音声とビデオのビットストリームを選択的に取り出し、音声とビデオの同期化のために適正な時刻表示が付加される。

【0026】図2と図3とから分かるように、従来の音声とビデオの符号化システムと異なる顕著な特徴は、音声ビットストリーム3およびビデオビットストリーム12が可変ビットレートのものであるということである。しかしながら、音声ビットストリーム3およびビデオビットストリーム12がシステム多重器14により同期化された

後の、出力システムビットストリーム15は、一定ビットレートである。

【0027】次に、可変ビットレート音声符号器について説明する。

【0028】ビット節約計数器と可変ビットレート音声符号器2との間の関係は、図4に具体的に示される。ビット節約計数器(BSC)25が上記に定義されるように、音声信号20は、毎秒 F_a フレームにセグメント化される。 i 番目の音声フレームのビット消費量は $f_a(i)$ ビット23であると仮定すると、1つの画像期間中のビット節約総数は、下記の(数1)により計算できることになる。

【0029】

【数1】

$$B_s = B_{smax} \cdot N_{af} - \sum_{i=1}^{N_{af}} f_a(i)$$

【0030】ただし、 N_{af} 28は、 B_s 要求がBAR 5からなされる前に符号化されている音声フレーム数である。 B_s 29がBAR 5へ一旦送られると、音声フレーム計数器27は、ゼロに設定され、次の画像期間中計数を継続する。

【0031】図5に図示されるように、音声フレーム毎にビット消費を求める目的のために、入力通話信号101は、予測分析モジュール102および音声の検出と識別のモジュール103へ渡される。予測分析モジュール102は、指定されたM階自己相関分析を使用する短期線形スペクトル分析を実施する。音声の検出と識別のモジュール103において、音声周波数帯域を通るエネルギー分布が計算される。無音がしきい値レベルに基づいて検出される場合、音声フレームの符号化は実施する必要はない。予測分析モジュール102から結果として生じる予測係数104と共に、音声セグメントは、有声または無音声成分(または一層小さい分類)への分類が実施される。用途の種類に応じて、各音声フレームの符号化に使用できる最大ビットレートは、先ずビットレート決定モジュール106において事前設定できる。予測係数とエネルギー分布情報から、使用できる最大ビットレートに等しいか、またはそれ未満の個別ビットレートレベルは、周波数重みつき信号対ノイズの測度、または音声フレームの符号化のための他の知覚測度に基づいて、ビットレート決定モジュール106において選定される。各音声フレームからのビット消費データ116は、ビット節約計数器へ伝送される。周期計数器から発生する音声フレーム番号115は、フレーム計数器へ伝送される。最大ビットレートよりも低いビットレートでの符号化は、コードブック108の適切な部分集合を使用することにより(もしくは減少されたコードブックにより、またはパラメータ的な表示方法用に再設計されたものにより)達成できる。多重コードブック108は、パルス状またはノイズ状のシー

ケンスを含む各種のシーケンスを発生できる励起発生器を構成する。励起シーケンスの合計109は、種々の予測器、例えばピッチ予測器やスペクトル予測器などから構成された予測器110に渡される。予測された信号111は、入力音声101から差し引かれて、差分すなわちエラー信号112を生じ、ついでこの信号は知覚的に重みづけられる。知覚的に重みづけされたエラー信号114は、ついでエラー最小化ベクトル量子化探索手順を駆動するのに使用される。ラインスペクトル対118および適切なパラメータ的な符号117は、音声ビットストリームの成分として伝送される。

【0032】上述の可変ビットレート動作は、従来のアルゴリズムとは異なるアプローチを有し、そのアルゴリズムは公称レートを有するが、公称レートより高いビットレートおよび低いビットレートで動作できる。

【0033】次に、ビットレート割り当て器(BRA)について説明する。図1に図示されるように、BRA 5は、音声符号器2から音声ビット節約総数 B_s 4を、ビデオ符号器10から先行符号化画像により使用されるビデオビット総数 B_v 7を、およびFIFOバッファの充足度 B_f 9を受信する。画像j用の事前設定目標ビット数を $B_t(j)$ とし、その $B_t(j)$ は、I画像符号化およびP画像符号化それぞれのための、事前設定I画像符号化ビットレート B_{tI} または事前設定P画像符号化ビットレート B_{tP} に等しいと仮定すると、j番目の画像符号化に利用できるビット総数は、下記の制御手段(数2)により計算されることになる。

【0034】

【数2】

$$\begin{aligned} B_a(j) &= B_t(j-1) - B_v(j-1) + B_s(j-1) + B_t(j) \\ \text{if } (B_a(j) + B_f(j) + B_{smax}) &> B_{fmax} \{ \\ &B_a(j) < B_{fmax} - B_f(j) - B_{smax} \\ &\} \\ \text{if } (B_a(j) + B_f(j) + B_{smax}) &< B_{Av} \{ \\ &B_a(j) > B_{Av} - B_f(j) - B_{smax} \\ &\} \end{aligned}$$

【0035】ただし、 B_{fmax} はFIFOバッファサイズであり、また B_{Av} は一定システムビットストリームビットレートである。j番目の画像符号化に利用できるビット数は、(j-1)番目の画像用の目標ビット数と(j-1)番目の画像により使用された実際のビット数との差に、(j-1)番目の画像の符号化期間中の音声符号器のビット節約数を加え、かつあらかじめ与えられたj番目の画像用の目標ビット数をさらに加えたものに等しい。利用できるビット数 $B_a(j)$ 、現行FIFOバッファ充足度 $B_f(j)$ および最大音声ビットレート B_{amax} の合計が、最大FIFOバッファサイズ B_{fmax} よりも大きいならば、画像jに利用できるビット数は、FIFOバッファ11のオーバーフローを避けるために、最大FIFOバッファサイズ B_{fmax} から現行FIFOバッファ充足度 $B_f(j)$ を差引き、また最大音声

ビットレート B_{amax} を差し引いたもの以下として計算されることになる。利用できるビット数 $B_a(j)$ 、現行 FIFO バッファ充足度 $B_f(j)$ および最大音声ビットレート B_{amax} の合計が、組合わせた音声とビデオのビットレート B_{AV} よりも小さいならば、画像 j に利用できるビット数は、FIFO バッファ 11 のアンダーフローを避けるために、 B_{AV} から現行 FIFO バッファ充足度 $B_f(j)$ を差し引き、また最大音声ビットレート B_{amax} を差し引いたもの以上として計算されることになる。

【0036】次に、可変ビットレートビデオ符号器について説明する。

【0037】可変ビットレートビデオ符号器 10 は、図 6 に図示されるように、符号化モード選定器 31 およびビット消費計数器 35 と共に構成される。現行画像に利用できるビット数 $B_a 30$ は、符号化モード選定器 (CMS) 31 へ送られる。 $B_a 30$ の値に応じて、CMS 31 は I 符号化、P 符号化または余分の P 画像符号化を選定するために意思決定プロセスを実施する。この決定結果は、現行画像 37 の符号化を制御するために可変ビットレートビデオ符号器 33 へ供給される。符号化されたビデオビットストリームは、リンク 34 を通して出力される。出力 34 はビット消費計数器 35 へにも送られるので、現行画像符号化に使用されるビット数 $B_u 36$ が計算されて、次の画像符号化に使用するために BAR 5 (図 1) へ送られる。

【0038】CMS 31 の意思決定プロセスは、図 7 のフローチャートで示される。I 画像の符号化に必要な最小ビット総数を B_{Imin} 、および P 画像の符号化に必要な最小ビット総数を B_{Pmin} と仮定する。さらに、先行の符号化された画像と、および先行の符号化された画像と符号化される現行画像との中間に位置する画像との間の画像相違 (PD) を検査するために事前設定されるしきい値 $TH1$ があると仮定する。

【0039】現行画像に利用できるビット総数 $B_a 50$ は、比較器 51 へ入力される。 B_a が B_{Pmin} の 2 倍よりも大きいならば、それは、利用できるビット総数は 2 つの P 画像を符号化するのに十分であることを意味する。言い換えれば余分の P 画像を挿入できることになる。この場合、比較器 51 の出力はリンク 53 を通り比較器 54 へ送られ、その比較器 54 は、先に定義した PD 値に基づいて P 画像を挿入する必要があるかどうかをさらに検査するのに使用される。PD 値が事前設定しきい値 $TH1$ よりも大きいならば、それは、先行の符号化された画像から、先行の符号化された画像と符号化される現行画像との中間に位置する画像まで、画像成分に実質的な変化があることを意味する。したがって決定プロセス 62 は、比較器 54 の出力がリンク 58 を通り「Yes」ならば、2 つの P 画像を符号化するように実施される。

【0040】しかしながら、比較器 54 がリンク 59 を通り「No」ならば、それは、先行の符号化された画像に関して実質的な変化が無く、また余分の P 画像を挿入する必

要がないことを意味する。利用できるビット総数 B_a は、ついで 1 つの I 画像の符号化または 1 つの P 画像の符号化の意思決定のために比較器 55 により、リンク 59 を通して検査される。このプロセスは、比較器 51 がリンク 52 から「No」を出力する場合にも適用される。したがって引き続き意思決定手順は、リンク 52 とリンク 59 との場合に適用される。

【0041】比較器 55 において、利用できるビット総数 B_a は、最小 I 画像符号化ビット総数 B_{Imin} と比較される。 $B_a \geq B_{Imin}$ ならば、それは、利用できるビット総数は 1 つの I 画像を符号化するのに十分であり、また決定プロセス 60 の I 画像符号化はリンク 56 から起動できることを意味する。 $B_a < B_{Imin}$ ならば、決定プロセス 61 の I 画像符号化は、リンク 57 からの出力により選定される。

【0042】60、61 および 62 の決定結果の 1 つだけが、一度に選定されて、63、64 および 65 のリンクの 1 つを通して符号化モード符号器 66 へ送られ、その符号器は、選定された決定結果を、出力リンク 67 に出力する。

【0043】上述した機構は、動的 I 画像および P 画像符号化モード制御と定義される。

【0044】画像符号化モードが決められると、 B_a は、現行画像符号化から発生するビットレートを動的に制御するように、量子化ステップを調整するのに使用される。この機構は、動的量子化ステップ制御と定義される。

【0045】ついで可変ビットレートビデオ符号器 33 は、CMS 31 により選定された画像符号化モードに従って現行画像を符号化するのに使用される。ビデオ符号器 33 の 1 つの好ましい実施例は、図 8 に示される変換符号化システムである。

【0046】各画像の画素データを含む入力ビデオ信号は、入力フレーム記憶装置 71 へ入力されて、格納される。ブロックサンプリング回路 72 は、入力フレーム記憶装置 71 に格納されたデータをライン 84 を通して受信し、その画像データを、画素データの空間的に非重複ブロックに分割する。妥当なレベルの適応性を得るために、 8×8 画素のブロックサイズを使用する。画像のフレーム番号は、入力画像記憶装置 71 へも入力され、またブロックサンプリング回路 72 を通りスイッチ (S1) 73 へ渡される。フレーム番号に基づいて、スイッチ 73 は、画像内符号化モードでの符号化のためにライン 86 を通して、または予測符号化モードでの符号化のためにライン 87 へ、ブロックサンプリング回路 72 から画素データの出力ブロックを選択的に送り出す。

【0047】画像内符号化モードの場合、ブロックサンプリング回路 72 の出力は、ライン 86 を通してディスクリート余弦変換 (DCT) 回路 76 へ渡される。DCT 回路 76 は、画像データを周波数領域におけるデータへ変換する周知の数学的変換であるディスクリート余弦変換を実施す

る。変換されたデータ、すなわちDCT 係数は、ついで量子化マトリックスと、およびレート制御装置79によりライン93を通して与えられる量子化ステップサイズとを使用する量子化回路77において量子化のプロセスを受ける。量子化されたデータは、補助情報としての量子化ステップと共に、ライン91を通り、ランレングス符号化および可変長符号化回路78へ渡され、その回路は、量子化されたデータおよびその関連する補助情報のランレングス符号化を実施し、ついでランレングス符号化結果の可変長符号化を実施する。ランレングス符号化および可変長符号化回路78の出力は、復号器への伝送が可能な符号化されたビットストリーム100 である。このビットストリームは又、ライン92を通してレート制御装置79へ渡される。ブロックの符号化のときに既に使用されたビット数に基づいて、レート制御装置79は、出力ビットストリームが符号化システムのビットレート必要条件を満足するように量子化ステップを調整する。量子化回路77により量子化された値は又、ライン94を通して、逆量子化回路80および逆ディスクリート余弦変換(逆DCT)回路81へ渡される。逆量子化回路80および逆DCT 回路は、DCT 回路76および量子化回路77により実施されるものの逆プロセスを実施して、再構成されたデータを得る。再構成されたデータは、ライン96を通して局所復号化画像記憶装置82に格納され、また予測符号化画像である次の入力画像のために動きの予測回路と補正回路に出力される。I 画像の全体が符号化された後に、レート制御装置は、I 画像の符号化に使用された合計ビットを計算して、次のP 画像およびI 画像に割り当てられるビット数を決定する。予測符号化モードの場合、ブロックサンプリング回路72の出力は、スイッチS1およびついでライン87を通して、動きベクトル(MV) 予測回路74へ渡される。MV予測回路74において、動きベクトル予測は、ライン98を通して得られる、局所復号化画像記憶装置82からの局所復号化画像を使用して、隣接画像からMVを決めるように実施される。MV予測回路74は、例えば、平均平方エラーのような幾つかの所定の規準に基づいて最良に一致したブロックを見出して、1つの画像から次の画像への画素のブロックの移行動き方向を求めることにより、最良の一致を見出す。

【0048】動きベクトル予測プロセスにおいて得られたMVは、符号化されるブロックと共に、ライン88を通り動き補正回路75へ渡され、その回路は、予測されたブロックを求める為に局所復号化画像記憶装置82に格納されたブロックを使用して動き補正を実施する。予測ブロックは、MVにより与えられるオフセットを局所復号化記憶装置から得られるブロックに適用して得られる。符号化されるブロックの画素の値と、予測ブロックの画素の値との間の差が計算されて、差分値から成る差分ブロックが得られる。差分ブロックに基づいて、動き補正回路75は又、ブロックの標準偏差などの幾つかの先に定義した

規準を使用して差分ブロックと符号化されるブロックの複雑さ(activity)を比較することにより、変換のためにDCT 回路76へ渡されるブロックを動き補正する必要があるかどうかを判断する。変換されるブロックが動き補正が必要であると判断されるならば、その差分ブロックはライン89を通してDCT 回路76へ渡される。変換されるブロックが動き補正する必要ないと判断されるならば、ブロック自体(すなわち、ブロックサンプリング回路72から出力された当初のブロック)は、ライン89を通してDCT 回路76へ渡される(それは、画像内符号化画像のブロックと同一である)。さらに動き補正回路75は、ブロックが動き補正で符号化されたかどうかを示す情報ビットを生成して、情報ビットを動きベクトルと共にDCT 回路76へ渡す。すなわち、ライン89を通して動き補正回路75からDCT 回路76へ出力されるデータは、差分ブロック(動き補正されて符号化された)または当初ブロック(動き補正無しで符号化された)であり、情報ビットは、動き補正されたか、否かの情報と差分ブロックの場合は動きベクトルの情報とを示す。

【0049】動き補正回路75からのブロックデータは、DCT 回路76においてDCT 係数へ変換され、DCT 係数は、量子化回路77において量子化されて、量子化DCT 係数となる。ブロックが動き補正またはMVで符号化されたかどうかを、それぞれ示す情報ビットは、DCT 回路76を通り、またライン90を通して量子化回路77へ渡される。量子化されたデータ(量子化されたDCT 係数)は、量子化ステップを含む補助情報と、動きベクトルと、および動き補正をブロックに実施したかどうかを、それぞれ示す情報ビットと共に、ランレングス符号化および可変長符号化回路78へ渡され、ランレングス符号化および可変長符号化されて、出力符号化ビットストリームとなる。そのビットストリームは又、ライン92を通してレート制御装置79へ渡される。量子化回路77からの量子化データは又、ライン94を通して渡され、逆量子化回路80において逆量子化され、逆DCT 回路81において逆DCT されて、次の画像の符号化のために局所符号化画像記憶装置82に格納される。

【0050】次に、FIFOバッファについて説明する。

【0051】図9 は、音声とビデオの符号化プロセス中におけるFIFOバッファ状態の一例を示す。第1の期間 T_d の間、符号化音声ビットストリーム3 は、FIFOバッファ11へ書き込まれる。瞬間 T_d から開始して、FIFOバッファ11は、音声ビットストリーム3 およびビデオビットストリーム12により充足される。FIFOバッファ11が事前設定充足度 B_0 まで充足されると、システム多重器14は、FIFOバッファ11からビットを収集することを開始する。システム多重器14が一定ビットレート B_{AV} でFIFOバッファ11からビットを取り出すので、バッファ充足度は、 B_{out} として示される、固定された総数に減少させられる。この値は常に一定である。音声とビデオのビッ

トレートは可変であるので、それらの組合わせも可変ビットレートである。したがってFIFOバッファ11のレートは、各期間におけるバッファ充足度カーブの異なる傾斜により図9で図示されるように、1つの期間から他の期間へ、変化している。B_{out}は、FIFOバッファ11のオーバーフローまたはアンダーフローを避けるように設定される。しかしながら、これも、図9の3T_dにより示されるように、伝送されるビットストリームに時間遅延を生じる。

【0052】次に、システム多重器について説明する。

【0053】システム多重器は、入力ビデオと音声データを固定長さのバケットにパケット化することにより、かつ適正な同期化を確実にするために、時刻表示を対応する音声フレームとビデオ画像に挿入することにより、FIFOバッファからの入力ビデオと音声データを多重化する。2つの時刻表示が使用される。すなわち、音声フレームまたはビデオ画像の提示ユニットを演奏または放送する必要がある時刻を示す提示時刻表示(PTS)、および音声またはビデオ画像を復号化する時刻を示す復号時刻表示(DTS)である。PTSおよびDTSは、システム

クロック基準(SCR)と呼ばれる共通時刻ベースを有して、測時の計測を統一し、かつ適正な同期化およびバッファ管理を確実にするようにする。

【0054】音声データの場合、システム多重器は、送られる画像数の追跡を続け、また新しいデータがFIFOバッファから得られると、適切なPTSがバケットヘッダーへ挿入される。DTSは、表示時刻が復号化時刻と固定関係を有するので、必要とされない。

【0055】ビデオデータの場合、システム多重器は、送られる画像数の追跡を続け、また適切なPTSおよびDTSがバケットヘッダーへ挿入される。DTSは、画像の復号化時刻と表示時刻が異なる場合には必要である。

【0056】

【発明の効果】以上述べたところから明らかなように、本発明は、画像符号化モードと共に、画像レートを制御する機構を提供することにより、固定された期間を通して利用できるビットを一層効率的に使用できるので、音声品質について妥協することなく、ビデオ信号の品質を大幅に向上できる。

【0057】本発明は、非常に低いビットレートのビデオと関連音声の符号化に特に効果的である。例えば、本発明がテレビ電話用途に適用される場合、しゃべっていない人の全ての音声ビットレートを利用することによりその人の、画像が、より多くのビットで符号化されて通話中の相手へ伝送されるようになり、画像品質を向上で

きる。

【図面の簡単な説明】

【図1】動的にビットレート制御されるビデオと関連音声の符号化システムのブロック線図を示す。

【図2】音声とビデオの符号化プロセス中の時間とビットレートの関係を図示する。

【図3】多重化プロセス中の時間とビットレートの関係を図示する。

【図4】ビット節約計数器と可変ビットレート音声符号器との間の関係を示す。

【図5】可変ビットレート音声符号器のブロック線図を示す。

【図6】符号化モード選定器およびビット消費計数器と共に、可変ビットレートビデオ符号器の構成のブロック線図を示す。

【図7】動的画像符号化モード選定フローチャートの機構を図示する。

【図8】ビデオ変換符号化方法のブロック線図を示す。

【図9】FIFOバッファ充足度を図示する。

【符号の説明】

- 1 音声信号
- 2 音声符号器
- 3 音声ビットストリーム
- 5 ビットレート割当器
- 6 ビデオ信号

- 10 ビデオ符号器
- 11 FIFOバッファ
- 12 ビデオビットストリーム
- 14 システム多重器
- 15 システムビットストリーム

B_u: 先行ビデオフレームにより使用されたビット総数
B_s: 現行ビデオフレームの前に音声フレームにより節約されたビット総数

B_f: 現行ビデオフレームの前のFIFOバッファ充足度

B_a: 現行フレームに利用できるビット数

B_{AV}: 組合わせられた音声とビデオのビットストリームに事前設定される固定ビット速度

B_{amax}: 最大音声ビットレート

B_{aavg}: 平均音声ビットレート

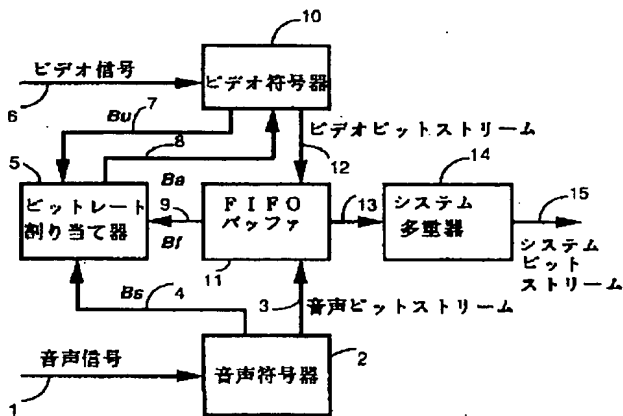
B_{vavg}: 平均ビデオビットレート

B'_{vavg}: 音声符号器により節約されたビットを付加した後の新しい平均ビデオビットレート

T_d: 音声符号器の開始点に対してのビデオ遅延時間

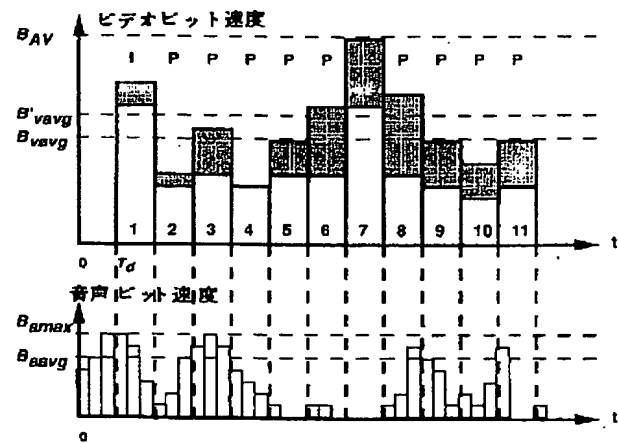
□ 音声符号器からのビット数

【図1】



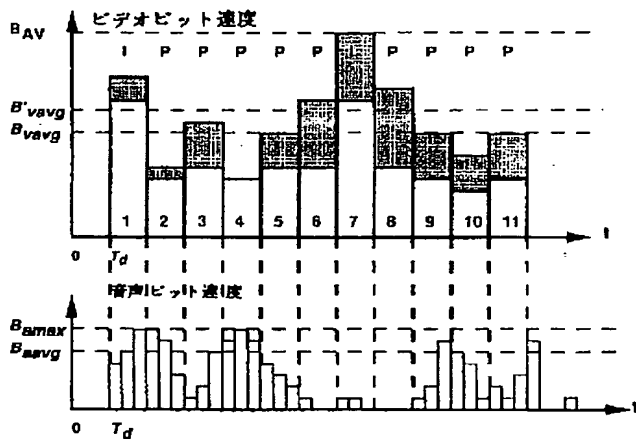
Bu: 先行ビデオフレームにより使用されたビット総数
 Bs: 現行ビデオフレームの前に音声フレームにより節約されたビット総数
 Bf: 現行ビデオフレームの前のFIFOバッファ充足度
 Ba: 現行フレームに利用できるビット数

【図2】



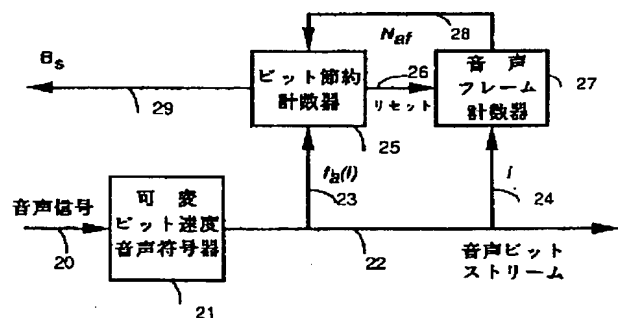
B_{AV} : 組合わせられた音声とビデオのビットストリームに事前設定される固定ビット速度
 B_{max} : 最大音声ビットレート
 B_{avg} : 平均音声ビットレート
 B'_{avg} : 平均ビデオビットレート
 B'_{avg} : 音声符号器により節約されたビットを付加した後の新しい平均ビデオビットレート
 T_d : 音声符号器の開始点に対してのビデオ遅延時間
 [Pattern]: 音声符号器からのビット数

【図3】

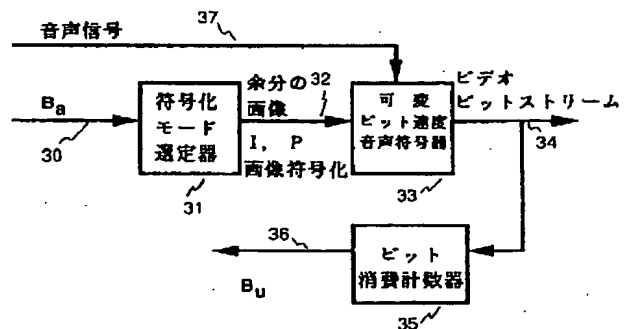


B_{AV} : 組合わせられた音声とビデオのビットストリームに事前設定される固定ビット速度
 B_{max} : 最大音声ビットレート
 B_{avg} : 平均音声ビットレート
 B'_{avg} : 平均ビデオビットレート
 B'_{avg} : 音声符号器により節約されたビットを付加した後の新しい平均ビデオビットレート
 T_d : 音声符号器の開始点に対してのビデオ遅延時間
 [Pattern]: 音声符号器からのビット数

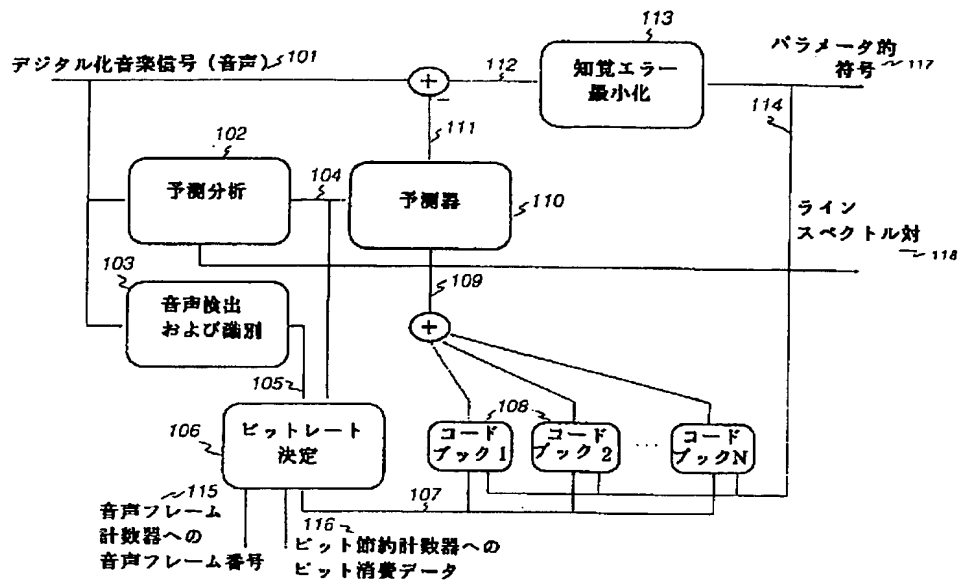
【図4】



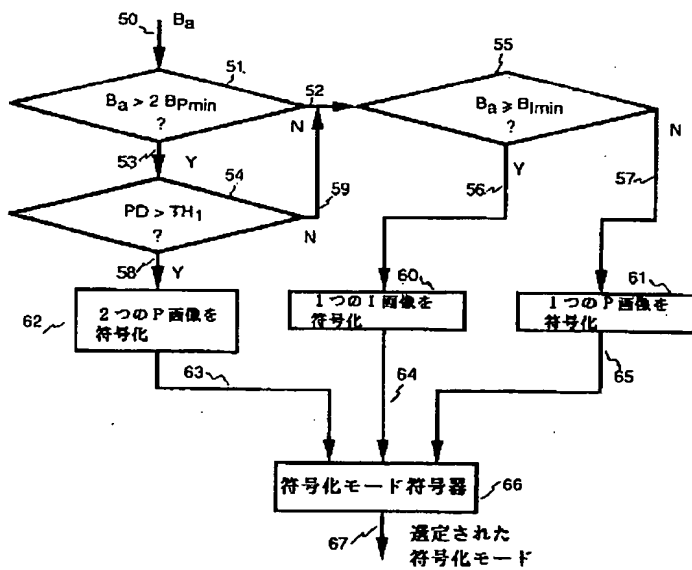
【図6】



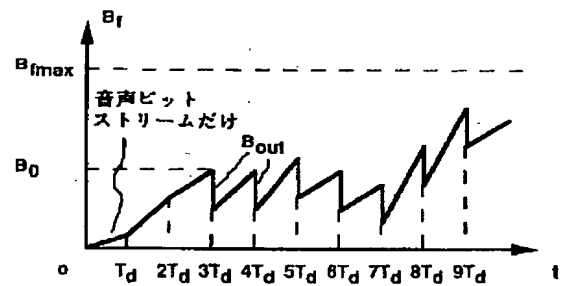
【図5】



【図7】



【図9】



【図 8】

